

*Uso inverso de las memorias de
traducción generadas a partir de la
traducción automática*

Nombre: Berta Marco Baeza

Línea de investigación: Tecnologías de la Traducción

Tutores: Felipe Sánchez Martínez y Mikel L. Forcada Zubizarreta

Fecha: 5 de junio de 2019

**Trabajo de
fin de grado de
Traducción e Interpretación**

Uso inverso de las memorias de traducción generadas a partir de la traducción automática

Berta Marco Baeza

bertamb17@alu.ua.es

RESUMEN

En este trabajo se estudia el uso inverso de las memorias de traducción generadas a partir de la traducción automática y se compara, por un lado, con el uso directo de una memoria de traducción profesional, que es el uso que más común en el trabajo de un traductor profesional y por otro, con el trabajo sin memoria de traducción. Para realizar este trabajo se ha colaborado con cuatro estudiantes de traducción e interpretación, quienes han traducido un texto (francés–español) con los tres métodos citados usando la herramienta de traducción asistida por ordenador OmegaT. Se les midió el tiempo de traducción en cada uno de los métodos y los resultados se compararon entre sí.

ABSTRACT

“Reverse use of translation memories generated from machine translation”

This paper studies the results of with the reverse use of translation memories generated from machine translation and compares them, on the one hand, with the results of a translation performed with the direct use of a professional translation memory, which is the most common in the a professional translator work and, on the other hand, without translation memory. To carry out this study we have worked with four students of translation and interpreting, who have translated text (french–spanish) with the three methods above using the

computer-aided translation tool OmegaT. Their translation time in each of the methods was measured and the results were compared.

Palabras clave: traducción automática, memoria de traducción, traducción inversa, herramienta de traducción asistida por ordenador.

Keywords: machine translation, memory translation, backtranslation, computer-aided translation tool.

ÍNDICE

1.	Introducción.....	1
2.	Definición de conceptos	2
2.1	Traducción automática	2
2.2	Programas de TAO basados en memorias de traducción	3
2.3	Las memorias de traducción	5
2.3.1	Usos de las memorias de traducción.....	6
2.3.2	Tipos de memorias de traducción	6
3.	Justificación del problema	9
4.	Objetivos.....	10
5.	Materiales y métodos.....	10
5.1	Elección de participantes	10
5.2	Creación del encargo de traducción.....	10
5.3	Creación de las memorias de traducción	12
5.4	Creación del glosario	13
5.5	Creación de los proyectos de traducción y reparto	14
5.6	Método experimental.....	15
6.	Resultados.....	16
6.1	Resultados detallados por fichero.....	17
6.2	Resultados por persona.....	19
6.3	Resultados del porcentaje de coincidencia medio relacionado con las estadísticas de coincidencia de la MTP y la MTA	19
7.	Discusión	21
8.	Conclusiones.....	22
9.	Bibliografía.....	23
10.	Anexo	27

1. Introducción

Las herramientas de traducción asistida por ordenador (TAO) están demostrando ser de gran utilidad en el trabajo profesional de traducción según muestran los estudios de traducción (Bermúdez, 2014: 130).

La mecanización de la traducción se hizo realidad en el siglo XX haciendo posible que se tradujeran una gran variedad de textos de un idioma a otro, pero esta práctica no está exenta de problemas. No obstante, el uso de traductores automáticos es un logro que ha facilitado muchísimo la labor de traducción profesional. Para evaluar esta ayuda lo que se debería estimar en realmente es la cantidad de postedición¹ que es necesaria hacer tras el uso de un traductor automático (Hutchins & Somers 1992: 2).

Una buena parte de los traductores profesionales trabajan para ofrecer una gran cantidad de documentos y textos de diversa índole cuyo contenido muchas veces es repetitivo, cuyo tiempo exigido de traducción normalmente es corto (Hutchins & Somers 1992: 2) y de cuyas traducciones se espera un trabajo de calidad. Por ese motivo, el empleo de herramientas de software específico de traducción, como son los programas de TAO, se convierte en un material de trabajo que puede ser útil y presenta muchas ventajas.

Los proveedores de tecnologías de la traducción a menudo afirman que la integración de su software en los procesos de traducción, ya sean sistemas de traducción automática (TA) o sistemas de memorias de traducción (MT) o ambos, pueden aumentar la productividad de los traductores. Las MT son am-

¹ La postedición consiste en corregir con los mínimos cambios posibles el resultado de una traducción automática para convertirla en una traducción adecuada para el propósito previsto.

pliamente utilizadas y sus usuarios afirman que son útiles en términos de productividad, consistencia o calidad (Bowker 2005: 19).

En este trabajo se quiere comprobar la eficiencia de un uso poco conocido de las MT, el uso inverso de las MT generadas a partir de la traducción automática (UIMTA), que se explicará más adelante.

Caso de uso del UIMTA:

Nos llega un encargo de traducción de un texto al que llamamos S' . Para traducirlo no tenemos una memoria de traducción profesional (MTP) útil que nos facilite el trabajo. En cambio, sí disponemos de un gran corpus de texto en LM al que llamamos T compuesto por textos de misma temática que los del encargo de traducción, pero en LM. Dicho corpus se compone de una cantidad de textos que se traduce automáticamente para obtener un corpus en LO al que llamamos Z ; mediante su alineación con el corpus T se crea la memoria de traducción automática (MTA).

2. Definición de conceptos

Conviene definir previamente los conceptos usados en este trabajo.

2.1 Traducción automática

La TA es el proceso de traducir un texto informatizado² de lengua origen a lengua meta haciendo uso de un programa de ordenador.

Una de las ventajas de usar un sistema de TA es que ello permite traducir grandes volúmenes de texto en un tiempo claramente inferior a la traducción profesional (Oliver 2007: 18). Los principales obstáculos a la hora de traducir

² Un texto informatizado es un fichero que contiene texto codificado de manera que puede ser modificado con un editor de texto. (Forcada et al. 2017: 97).

con un sistema de TA son los de tipo lingüístico, entre los que se encuentran los problemas de ambigüedad léxica, complejidad sintáctica, diferencias para referirse a un mismo concepto entre lenguas, etc. (Arnold 2003). Estas limitaciones comportan una serie de recomendaciones para que los traductores automáticos sean rentables: la principal es el estilo del texto. Se recomienda que el texto que vaya a ser traducido con un traductor automático esté escrito con frases cortas, con pocos pronombres y con poca complejidad sintáctica, es decir, en un lenguaje controlado (Oliver 2007: 16–19). El lenguaje controlado está diseñado de forma que el resultado de la traducción automática pueda ser usado directamente con el mínimo posible de postedición (Forcada et al. 2017: 108). Sin embargo, como la mayoría de los textos originales no se escriben teniendo en cuenta este tipo de lenguaje no podemos esperar esto.

2.2 Programas de TAO basados en memorias de traducción

Los sistemas de TAO son un conjunto de programas informáticos que ayudan al profesional de la traducción a traducir textos de forma rápida, ágil y le ayudan a hacerlo con un alto nivel de calidad (Oliver 2007: 28). Es decir, el término TAO (en inglés *computer-aided translation*; *CAT*) se usa normalmente para referirse al software que permite la traducción profesional con apoyo de la TA, de las sugerencias de traducción de una memoria de traducción (MT) o de bases de datos terminológicas (Forcada et al. 2017: 106). Por tanto, la TA, las sugerencias de una memoria de traducción o las bases de datos terminológicas por sí solas no son herramientas de TAO, sino componentes de estas.

Una de las ventajas de usar una herramienta de TAO es que esta incrementa la productividad gracias a la reutilización de las unidades de traducción (UT)

obtenidas a partir de traducciones anteriores (Cámara 2001). Por eso, las razones del empleo de este software aparecen sobre todo cuando los textos son repetitivos, de temática similar o cuando se necesitan introducir algunos cambios (por ejemplo, en la actualización de las traducciones de publicaciones técnicas, normativas, reglamentos, etc.), ya que se pueden encontrar segmentos similares de otras traducciones (López 2003: 108)

¿Cómo funciona un programa de TAO?

Un programa de TAO basado en MT organiza las UT en una base de datos que permite recuperarlas cuando se va a traducir un texto nuevo. Estas UT se van almacenando y se pueden crear de dos formas: o bien a medida que un traductor profesional va traduciendo nuevos textos usando la propia herramienta de TAO o bien al alinear nosotros mismos segmentos originales y traducciones existentes para generar un conjunto de UT. La similitud entre una nueva oración s' y la parte origen s de una UT (s, t) se calcula con el porcentaje de coincidencia³ (en inglés *Fuzzy Match Score: FMS*). Dependiendo de este porcentaje de coincidencia se habla de coincidencia exacta o *exact match* cuando el sistema encuentra un segmento idéntico, es decir, del 100%. Por otra parte, se habla de coincidencia parcial o *fuzzy match* cuando el sistema identifica una UT con un porcentaje de coincidencia inferior al 100% (Forcada, et al. 2014: 42).

³ El porcentaje de coincidencia es el porcentaje entre el segmento origen y un segmento de memoria de traducción. (Forcada et al. 2017: 217).

Esta es la ecuación con la que se calcula el porcentaje de coincidencia.

$$\text{FMS}(s, s') = 100\% \cdot \left(1 - \frac{\text{ED}(s, s')}{\max(|s'|, |s|)} \right)$$

El porcentaje de coincidencia entre s y s' se calcula dividiendo la ED (*edit distance* en inglés), es decir, la distancia de edición entre el segmento s' y s dividido entre el máximo de palabras de ambos segmentos. La distancia de edición es el número mínimo de cambios (eliminaciones, inserciones y sustituciones) necesarios para transformar s en s' (Ortega et al. 2014: 42). Por tanto, cuanto mayor es ED(s, s') menor es el FMS.

Según Bowker (2002: 100) en los programas de TAO el usuario puede establecer un umbral de coincidencia para que el sistema presente sugerencias a partir de ese porcentaje. Dicho umbral se suele establecer por encima del 60% (Esplà-Gomis et al. 2011: 177)

2.3 Las memorias de traducción

Una MT es un repositorio de segmentos⁴ de texto en LO que deben estar segmentados para alinearse con sus correspondientes segmentos traducidos en LM para obtener como resultado una lista de UT (s, t) donde s es el segmento original y t el segmento destino (Oliver 2007: 46-47). Para ello se pueden usar alineadores de textos como YouAlign,⁵ aunque las últimas actualizaciones de

⁴Los segmentos de las MT normalmente se corresponden con oraciones, aunque también se pueden segmentar en párrafos.

⁵<https://youalign.com/>

los sistemas de TAO basados en memorias de traducción como OmegaT⁶ o SDL Trados Studio⁷ llevan incluido un alineador, por lo que este último paso podría realizarse en el mismo programa.

Antes de entrar a explicar los tipos y usos de las MT, conviene remarcar un apunte para comprender los términos que se emplearán a lo largo del trabajo.

Un texto puede ser traducido en dos direcciones, la directa y la inversa. La traducción directa L_2-L_1 ⁸ es aquella en la que un texto en lengua L_2 se traduce a la L_1 . Por el contrario, la traducción inversa L_1-L_2 es aquella en la que un texto en L_1 se traduce a la L_2 . Por último, cabe remarcar que estos conceptos de «directa» e «inversa» no son los que se usan en este trabajo.

A continuación, explicaremos los usos y los tipos de las MT (ver tabla 1).

2.3.1 Usos de las memorias de traducción

En un encargo de traducción en el que se trabaje con una MT, independientemente de que sea traducción directa o inversa, esta se puede usar de dos maneras diferentes: uso directo (UDMT) o un uso inverso (UIMT). Conviene recalcar que hacer un uso inverso de una MT no significa que la traducción en la que se esté trabajando tenga que ser inversa.

2.3.2 Tipos de memorias de traducción

Dependiendo de quién haya traducido el texto que ha generado las MT podemos hablar de dos tipos de MT:

⁶<http://omegat.org/>

⁷<https://www.sdltrados.com/es/>

⁸ Hablamos de L_1 y L_2 para referirnos a las lenguas con las que trabaja un traductor profesional, siendo la L_1 la de mayor dominio.

▪ *Memoria de traducción profesional*

Cuando hablamos de MTP nos referimos a que dicha MT se ha generado a partir de traducciones profesionales, independientemente de que hayan sido traducciones directas o inversas. Sin embargo, lo que sí se puede diferenciar es el uso que se haga de la MT. En este sentido, distinguiremos el uso directo (UDMTP), que se empleó en este trabajo, y el uso inverso (UIMTP).

Cabe mencionar que posiblemente muchas de las MTP se hayan creado a partir de una traducción directa porque, como se ha dicho antes, este es el sentido en el que los traductores profesionales suelen trabajar. Esto quiere decir que, si un traductor profesional (cuya L_I es el español) recibe un encargo de traducción directa con el par de lenguas y lo traduce haciendo uso de un programa de TAO con MT memorias generadas a partir de una traducción directa, estará haciendo un uso directo UDMTP.

En este sentido, el UDMTP es el que más abunda en el sector, ya que si, por el contrario, un profesional recibiera un encargo de traducción inversa posiblemente buscaría MT generadas a partir de una traducción directa, aunque creadas con el par de lenguas inverso (francés–español). No obstante, esto no quita que a veces se tenga que optar por el uso inverso de una memoria de traducción profesional (UIMTP), al no disponer de otra.

En el caso de la MTP, que se descargó de la página web OPUS,⁹ no se puede garantizar que las UT sean traducciones naturales o profesionales unas de otras. Es decir, lo más probable es que esta MT se haya generado usando como lengua pivote el inglés, de manera que haya habido una traducción LO–LM inglés–español, otra traducción LO–LM inglés–francés y ambas LM francés–español se hayan alineado.

⁹ <http://opus.nlpl.eu/EMEA.php>

Sin embargo, se puede crear otro tipo de MT usando un sistema de TA: las *memorias de traducción automática* (MTA).

▪ *Memoria de traducción automática*

Cuando hablamos de MTA nos referimos a que dicha MT se ha creado a partir de un texto natural en LO que ha sido traducido con un sistema de TA para obtener un texto en LM, independientemente de si ha sido una traducción directa o inversa.

Como en las MTP, lo que sí se va a diferenciar es el uso que se haga de ella, que puede ser el directo (UDMTA) o el inverso (UIMTA); este último es el objeto de este trabajo. En UDMTA, en cada UT (s, t) t es un segmento artificial traducción automática de s , que es un segmento natural. Sin embargo, en UIMTA, en cada UT (z, t) z es un segmento artificial producido por la traducción automática de un segmento natural t . En UIMTA, los segmentos z de las UT (z, t) son una aproximación de los segmentos naturales s que serían a su vez equivalentes a los segmentos t de las UT (s, t) . La naturaleza de la concordancia es, por tanto, muy diferente a la que se produce en otros usos de las MT, en las que la concordancia se establece estos segmentos naturales.

Tabla 1: Tipos y usos de las memorias de traducción.

		Uso de la MT	
		directo (UD)	inverso (UI)
Tipo de memoria de traducción	profesional (MTP)	UDMTP	UIMTP
	automática (MTA)	UDMTA	UIMTA

El objetivo de este trabajo es comparar la eficiencia del UIMTA, ya que es poco conocido y probablemente poco usado con lo que asumimos que es UDMTP (uso directo) que como se ha dicho antes, es el uso más común de una MT.

Es importante tener en cuenta que la calidad de la MTA dependerá de la calidad del sistema de TA.

3. Justificación del problema

Los estudios de Koehn (2009), Brkić et al. (2009), Vázquez et al. (2013), Screen (2016) y Screen (2017) en los que se midió la productividad de los traductores a la hora de traducir (hacia su L_I) con herramientas de TAO y sin ninguna herramienta concluyeron que los grupos que tradujeron con ayuda de MT convencional o TA aumentaban su productividad con respecto al grupo que tradujo sin ninguna ayuda. Sin embargo, hasta la fecha de hoy no conocemos ni estudios ni trabajos ni artículos que en los que se hable del UIMTA, pero... ¿Es posible que este uso inverso de la MTA sea igual o más eficiente que el uso directo de las MTP? Despejar esta duda es el principal motivo de este trabajo.

Por eso este trabajo se quiere comprobar si el nuevo método de traducir con el UIMTA es igual de eficaz o no con respecto al UDMTP. Para ello, se usarán las tres herramientas anteriormente citadas: la TA para crear la MTA, las MTP y MTA, y la herramienta de TAO para realizar las traducciones.

4. Objetivos

Estudiar la eficiencia de la traducción mediante UIMTA y comparar los resultados con los de la traducción con UDMTP y con los obtenidos usando ningún recurso; finalmente sacar conclusiones.

5. Materiales y métodos

5.1 Elección de participantes

Se trabajó con un grupo de cuatro estudiantes (*P1*, *P2*, *P3*, *P4*) del grado de traducción e interpretación (francés) de la Universidad de Alicante de último curso y recién graduados, cuya L_1 es el español.

5.2 Creación del encargo de traducción

Para realizar este trabajo, se parte de la base de que a los participantes les ha llegado un encargo de traducción francés–español al que llamamos *S'*, formado por tres dictámenes de la Agencia Europea del Medicamento (EMA), en concreto los dictámenes del medicamento *Jinarc*, *Kyprolis* y *Tremfya*. Además, se seleccionó el dictamen de otro medicamento *Memantine Mylan*, con el que se entrenaron antes de empezar a traducir.

Es cierto que en la vida real el encargo de traducción *S'* no se elige, pero en el caso de este trabajo, al tratarse de una situación artificial, se seleccionaron los textos por una razón. El motivo de elección fue la similitud del estilo en el que estaban escritos, la estructura similar que presentaban y el uso del lenguaje propio de esa tipología textual entre ellos y con los textos que forman las MT que se usarán.

Los textos que componían el encargo de traducción *S'* y el corpus *T* solamente estaban disponibles en formato PDF. Como la traducción se realizó con

el sistema TAO basado en memorias de traducción OmegaT (versión 3.4) que no permite documentos en formato PDF,¹⁰ fue necesario convertir dichos textos en documentos de texto sin formato (.txt) con el editor de textos Notepad++,¹¹ editor con el que también se segmentaron. Después se juntaron en un solo documento y, al ver que el encargo era demasiado largo para llevar a cabo el experimento, cada texto que lo componía se redujo en palabras (ver tabla 2).

Tabla 2: Longitud de los textos del encargo de traducción.

Formación del encargo		
	N.º de palabras inicial	N.º de palabras final
Memantine Mylan *	888	566
Jinarc	1229	711
Kyprolis	1052	756
Tremfya	909	709
Total	3190	2158

** Este texto no entra en la suma total de palabras traducidas.*

¹⁰ Casi ningún sistema de TAO admite documentos en formato PDF.

¹¹ <https://notepad-plus-plus.org/>

5.3 Creación de las memorias de traducción

En segundo lugar, se prepararon las MT con las que se iba a traducir.

Por una parte se descargó en formato TMX¹² la MTP de la página web OPUS, subcorpus de EMEA, que está formada por el par de lenguas francés–español y tiene un total de 394.253 UT. Tras analizarla con el programa OmegaT junto con el texto del encargo *S'* presentaba las siguientes coincidencias para todos los textos en total (ver tabla 3).

Por otra parte, se creó la MTA. Esta memoria se creó a partir de un corpus *T* de temática similar a los del encargo (los segmentos en LM en español de la MTP) y se tradujeron con el sistema de TA Apertium¹³ que se instaló en un ordenador para crear el corpus de texto *Z* en francés. La MTA tiene un total de 317.998 UT. Por último, tras alinear ambos corpus, se obtuvo la MTA. Tras analizarla con el programa OmegaT presentaba los siguientes porcentajes de coincidencia (ver tabla 3).

¹² TMX es el formato estándar en el que se intercambian las MT, que está basado en XML (Forcada et al. 2017:217).

¹³ <https://www.apertium.org/index.spa.html?dir=spa-fra#translation>

Tabla 3: Tabla que muestra las coincidencias de MTP y MTA para todos los textos en total.

% de coincidencia entre la MTP y S'			% de coincidencia entre la MTA y S'		
	N.º segmentos	N.º palabras		N.º segmentos	N.º palabras
Repeticiones	2	22	Repeticiones	2	22
Coincidencia exacta	0	0	Coincidencia exacta	0	0
95%-100%	18	263	95%-100%	0	0
85%-94%	32	425	85%-94%	4	53
75%-84%	28	377	75%-84%	11	136
50%-74%	105	2250	50%-74%	110	1830
Sin coincidencia	82	2354	Sin coincidencia	140	3650
Total:	267	5691	Total:	267	5691

5.4 Creación del glosario

Por último, también se creó un pequeño glosario de términos a partir de los textos del encargo que se incluyó en el programa de TAO y sirvió de apoyo para traducir. Como los textos del EMEA se pueden encontrar en la página web en ambos idiomas francés y español) se descargaron en los dos idiomas. Para crear el glosario primero se hizo una lectura exhaustiva de los tres dictámenes en español y se fueron subrayando los términos técnicos y colocaciones. Solo fue cuestión de anotar los términos en español y buscar la traducción del término en cuestión en el texto en francés. Finalmente, el glosario se elaboró con el editor de textos Notepad++, el formato requerido por OmegaT.

5.5 Creación de los proyectos de traducción y reparto

Una vez creados el encargo de traducción, las dos MT y el glosario se crearon los proyectos de traducción con el programa de TAO OmegaT (versión 4.1.1).

A cada uno de los participantes les llegó un encargo de traducción (francés español) de un texto S' compuesto por tres textos. Los participantes tradujeron cada uno de los textos de un modo diferente (UDMTP, UIMTA y nada). En total se crearon 12 proyectos más uno de entrenamiento que fue el mismo para todos los participantes, que se repartieron de la siguiente manera (ver tabla 4).

Siempre que se crea un proyecto de traducción es necesario declarar cuáles son las lenguas de trabajo; en este caso fueron francés–español. Todos los proyectos de traducción contienen las siguientes carpetas: *dictionary*, *glossary*, *OmegaT*, *source*, *target* y *tm*. En este trabajo solo se empleó la carpeta *glossary*, en la que se introdujo el glosario, la carpeta *source*, en la que se metió el encargo de traducción S' y la carpeta *tm*, en la que se situó la MTP en los proyectos en los que se hizo uso de esa memoria y la MTA en los que se hizo uso de esta otra. En la carpeta *dictionary* no se situó nada y en la carpeta *target* es donde OmegaT va guardando la traducción realizada.

Tabla 4: Asignación de los textos y modo de traducción para los participantes.

Reparto de proyectos				
	P1	P2	P3	P4
Memantine Mylan	entrenamiento	entrenamiento	entrenamiento	entrenamiento
Jinarc	nada	UIMTA	UIMTA	UDMTP
Kyprolis	UDMTP	nada	UDMTP	UIMTA
Tremfya	UIMTA	UDMTP	nada	nada

Una vez realizados estos pasos ya se contaba con todo lo necesario para poder empezar a traducir.

5.6 Método experimental

Se quedó con los participantes para traducir en días diferentes por cuestiones de tiempo. Se les dijo a los participantes que los encargos venían de la Agencia Europea del Medicamento.

Antes de comenzar con la traducción de los textos los participantes tradujeron el texto de entrenamiento que sirvió tanto de calentamiento como para familiarizarse con el programa TAO.

Los cuatro participantes utilizaron para traducir la misma versión del programa TAO OmegaT (versión 3.5) a la que se le incluyó una extensión para cronometrar el tiempo. Dicha extensión es «silenciosa» y no interfiere en la traducción. Todos ellos tradujeron en su zona de confort (zona donde suelen trabajar de una manera cómoda) con el mismo ordenador, un portátil Toshiba de 4 años de antigüedad.

En cada uno de los tres proyectos a los participantes se les presentó para traducir un texto sin MT, otro texto con MTP y otro texto con MTA, pero no se les dijo qué MT se ofrecía en cada texto.

Cada uno de los participantes tradujo alrededor de 2000 palabras en total. No había límite de tiempo para acabar las traducciones. Además, los participantes pudieron consultar en internet fuentes externas para ayudarse en la traducción.

6. Resultados

La tabla 5 muestra los datos obtenidos (la media) de velocidad de traducción por texto, participante y modo de traducción y en la figura 1 se pueden ver estos resultados de forma visual.

Primero comentaremos los resultados globales tomando de referencia la V_2 (palabras/segundo), cuando se ha traducido sin ayuda, con UDMTP y con UIMTA. En segundo lugar, se comentarán los resultados en función de si ha habido coincidencia o no con la MT (ver tabla 5); para ello se ha calculado, por una parte, V_3 (palabras/segundo) solo para los segmentos en los que la MT sí ofrecía coincidencias y, por otra parte, en los que no la ofrecía V_4 (palabras/segundo). De esta manera se puede comprobar cuál de ellas beneficia más al traductor.

Por último, se comparan los resultados del porcentaje de coincidencia medio relacionándolo con las estadísticas de coincidencia de la MTP y la MTA (ver tabla 5) y la calidad del TA.

Tabla 5: Tabla de tiempos de cada traductor agrupados por los textos del encargo.

Medias							
		Tiempo	V ₁	V ₂	V ₃ Sí MT	V ₄ No MT	% coincidencia medio
		Seg/segmento	Caracter/seg	Pal/seg	Pal/seg	Pal/seg	
Jinarc	P1-nada	69,45	1,74	0,28	-	0,28	-
	P2-UIMTA	50,19	2,40	0,38	0,35	0,40	11,78 %
	P3-UIMTA	63,24	1,91	0,30	0,29	0,31	11,78 %
	P4-UDMTP	77,02	1,74	0,28	0,26	0,31	27,49 %
	Medias	64,98	1,95	0,31	0,30	0,33	17,02 %
Kyprolis	P1-UDMTP	98,15	1,41	0,21	0,22	0,19	33,07 %
	P2-nada	93,62	1,48	0,22	-	0,22	-
	P3-UDMTP	57,39	2,61	0,40	0,41	0,36	33,07 %
	P4-UIMTA	67,93	2,03	0,31	0,38	0,28	13,94 %
	Medias	79,27	1,88	0,29	0,34	0,26	26,69 %
Tremfya	P1-UIMTA	61,48	1,94	0,30	0,30	0,31	20,33 %
	P2-UDMTP	58,38	2,05	0,32	0,32	0,30	36,01 %
	P3-nada	66,43	1,77	0,28	-	0,28	-
	P4-nada	53,24	2,31	0,36	-	0,36	-
	Medias	59,88	2,02	0,32	0,31	0,31	28,17 %

6.1 Resultados detallados por fichero

Jinarc: P2 y P3 (UIMTA) han sido más veloces que P4 (UDMTP). Por otra parte, P4 y P1 (nada) son igualmente de veloces, por lo que deducimos que en este texto el UIMTA ha beneficiado al traductor. En función de si hay coincidencia o no, por lo que observamos en P2, P3 y P4, en los segmentos que sí mostraban UT la velocidad fue menor en comparación con los segmentos que no la mostraban, por lo que cuando el programa TAO no propuso una UT, la

velocidad incrementó. Sin embargo, fijándonos solo en los segmentos en los que sí se mostraba una UT la velocidad fue mayor en los participantes que tradujeron con UIMTA (P2 y P3), aun siendo menor el porcentaje de coincidencia medio en comparación con la UMTPD.

Kyprolis: P3 (UDMTP) tiene mayor velocidad de traducción, seguido de P4 (UIMTA), de P2 (nada) y de P1 (UDMTP). Sin embargo, P3 (UDMTP) es más veloz que P4 (UIMTA), por lo que podemos deducir que para unos sí beneficia y para otros no (habilidad individual de cada traductor con la herramienta). En función de si hay coincidencia o no, para P1, P3 y P4, en los segmentos que el programa TAO sí proponía una UT la velocidad fue mayor en comparación con los segmentos que no la mostraban. Sin embargo, fijándonos solo en los segmentos en los que sí se mostraba una UT, la velocidad mayor fue la de P3 (UDMTP), seguido de P4 (UIMTA) y de P1 (UDMTP). Paralelamente, fijándonos en los porcentajes de coincidencia, observamos que los participantes que tradujeron con UDMTP es mayor que la del participante que tradujo con UIMTA.

Tremfya: P4 (nada) tiene mayor velocidad de traducción, seguido de P2 (UDMTP), de P1 (UIMTA) y de P3 (nada). Si nos fijamos solo en los participantes que sí usaron MT, vemos que P1 (UDMTP) fue un poco más veloz que P2 (UIMTA), teniendo el doble de porcentaje de coincidencia de las MT. Fijándonos en los datos cuando el programa sí mostraba UT, vemos que P2 (UDMTP) ha sido más veloz que P1 (UIMTA).

6.2 Resultados por persona

A continuación, en la tabla 6 se muestran los modos que ha usado cada participante ordenados de mayor a menor velocidad de traducción.

Tabla 6: Tabla que muestra las velocidades de mayor a menor por traductor.

Resultados por persona V ₂	
Participante	V ₂ según los modos de traducción
P1	UIMTA> nada> UDMTP
P2	UIMTA> UDMT> nada
P3	UDMTP> UIMTA>nada
P4	UIMTA> nada> UDMTP

6.3 Resultados del porcentaje de coincidencia medio relacionado con las estadísticas de coincidencia de la MTP y la MTA

Independientemente de los textos del encargo de traducción, P1, P2 y P3, cuando tradujeron con la MTP tradujeron más rápido los segmentos para los que sí se mostraban coincidencias que para lo que no. Sin embargo, cuando usaron la MTA, tradujeron más despacio los segmentos para los que sí se mostraban coincidencias que para los que no. Esto puede deberse a que el número de coincidencias de la MTP de los porcentajes comprendidos entre el 75% y el 100% es mayor y a la diferencia media del porcentaje de coincidencia de la MTP y MTA, que es de 17% aproximadamente teniendo mayor coincidencia la MTP.

Figura 1

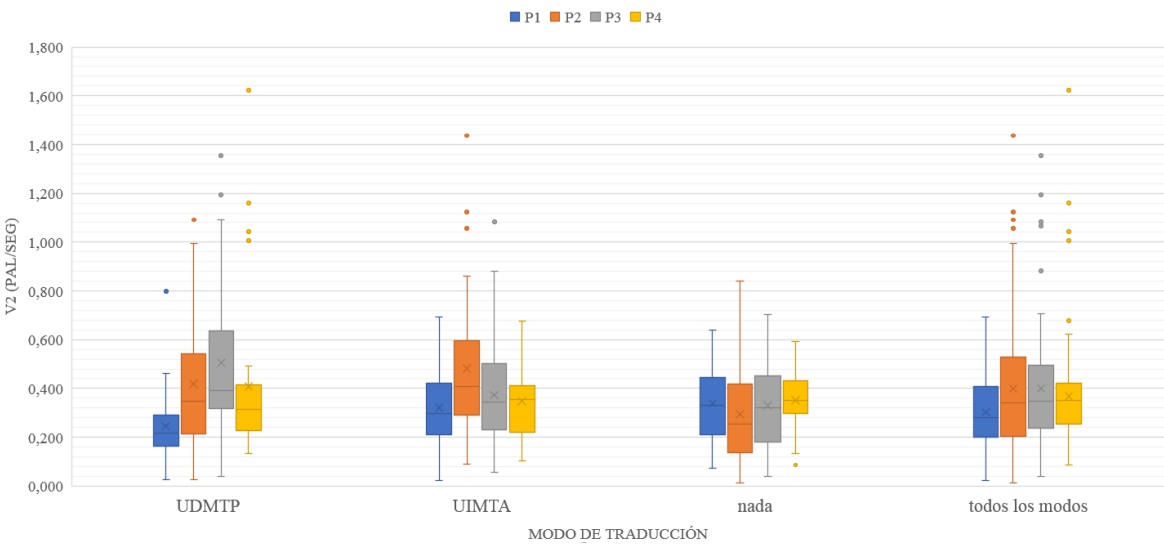


Figura 1: Gráfico de cajas y bigotes en el que se muestra la velocidad de traducción de cada participante en cada modo de traducción y con todos los modos juntos.

Si observamos la figura 1 podemos ver que en UDMTP ha ayudado más a P2 y P3. El UIMTA también ha beneficiado a P2 y P3, pero la diferencia de velocidad con respecto a P1 y P4 se recorta. Sin embargo, la velocidad de los cuatro traductores cuando han traducido sin ningún recurso se iguala. Por último, si nos fijamos en el cuarto bloque de cubos que representa la suma de las velocidades de cada traductor con los tres modos juntos vemos que, P2 y P3 han traducido más rápido pero no se despegan tanto de P1 y P4, los más lentos.

7. Discusión

Como se puede comprobar, la productividad de los textos traducidos con el UIMTA no ha sido mayor en comparación con UDMTP. En primer lugar, ello se debe a que no había tantas coincidencias como en el texto con la ayuda de la MTPD. Sí que es cierto que el porcentaje de coincidencias de la MTAI era mayor, es decir, las unidades de traducción que ofrecía el programa de TAO eran más útiles, pero estas coincidencias aparecían la mitad de las veces que en UDMTP y para segmentos más cortos. Esto puede deberse a la calidad del TA, que no es tan bueno traduciendo de francés a español.

Para próximos estudios de este estilo se recomienda que, a pesar de contar con un cronómetro, se cronometre aparte el tiempo de traducción, ya que se corre el riesgo de que pueda fallar la extensión experimental.

Limitaciones del trabajo

A lo largo de la realización del trabajo hubo algunos problemas que ralentizaron su ejecución. Uno de los problemas con los que nos encontramos fue que la extensión experimental (*session-log*),¹⁴ que cronometró el tiempo de traducción por segmento, en dos traducciones no lo hizo, por lo que no se obtuvieron datos de tiempo de esos dos textos. Esto hizo que dos de los participantes (P1-*Tremfya* y P2-*Jinarc*) tuvieran que repetir de nuevo la traducción y puede haber ocasionado alteraciones en los resultados. Si observamos los datos de P2-*Jinarc* vemos que ha sido la más rápida en traducir dicho texto; esto ha podido deberse a que tuviera sesgo de recuerdo del texto. Por el contrario, en P1-*Tremfya* no se ha dado este resultado.

¹⁴ <https://github.com/mespla/OmegaT-SessionLog>

Dicha extensión experimental no era compatible con la versión 4.1.4 de OmegaT que era la actual en el momento de los estudios, por ello los experimentos se tradujeron con la versión 3.5, que sí era compatible.

Ambas MT eran tan grandes que en ocasiones el programa de TAO OmegaT tardó en ofrecer las coincidencias por lo que, en algunos segmentos de un mismo texto, dependiendo del participante, si este no se esperó lo suficiente a que el programa de TAO le mostrara las coincidencias, pudo no haberlas visto. Esto también puede haber ocasionado desviación en los resultados.

8. Conclusiones

Los resultados obtenidos en este estudio en cuanto al tiempo medio de traducción sin ayuda, con UDMTP y con UIMTA, tal y como se observa en la figura anterior, son bastante similares entre sí y por tanto no son concluyentes. Fijándonos en el diagrama de barras y bigotes. Podemos observar que, en general, la variación de velocidad entre los tres modos no es muy grande, por lo que no se puede decir las MT hayan sido de gran ayuda. Sin embargo, podemos distinguir dos tipos de traductor: para P2 y P3 parece ser que sí ha habido un ligero aumento de velocidad con el UDMTP y el UIMTA.

Bien es cierto que al no haber trabajado con más participantes y no tener con más datos, estos resultados hay que tomarlos con cautela. Por este motivo no podemos afirmar que un uso sea más eficiente que el otro.

9. Bibliografia

- Arnold, Doug. (2003) “Why translation is difficult for computers.” En: Somers, Harold. (ed.), *Computers and Translation: A translator’s guide*. Amsterdam y Philadelphia: John Benjamins, pp.119–142.
- Bowker, L. (2005). “Productivity vs quality? A pilot study on the impact of translation memory systems.” *Localisation Focus* 4:1, pp.13–20. Versión electrónica <
https://www.localisation.ie/sites/default/files/publications/Vol4_1Bowker.pdf>
- Bowker, Lynne (2002) *Computer-aided translation technology: a practical introduction*, chapter Translation-Memory Systems, pages 92–127. University of Ottawa Press.
- Brkić, Marija; Sanja Seljan & Božena Bašić Mikulić. (2009) “Using Translation Memory to speed up translation process.” En: Bawden, David; Damir Boras & Senada Dizdar (eds.) 2009. 2nd International Conference. “*The Future of Information Sciences: INFUTURE2009: Digital Resources and Knowledge Sharing*.” Zagreb: Editorial, pp. 353–363. Versión electrónica <
<http://infoz.ffzg.hr/INFUTURE/2009/papers/INFUTURE2009.pdf>>
- Cámara, Lidia. (2001) “El papel de las herramientas TAO en la documentación técnica multilingüe” *Revista Tradumàtica: traducció i tecnologies de la informació i la comunicació* V:0. Versión electrónica
<<http://www.fti.uab.es/tradumatica/revista/num0/mapa/mapa.htm>
- Esplà Gomis, Miquel; Felipe Sánchez Martínez & Mikel L. Forcada. (2011) “Using machine translation in computer-aided translation to suggest the target-side words to change.” *Transducens comunicacions y congresos*, pp.172–178. Versión electrónica <
<http://rua.ua.es/dspace/handle/10045/27578>>
- Forcada, M.L., Sánchez Martínez, F., Pérez Ortiz, J.A. (2017) *Manual de informática y tecnologías para la traducción*. Universitat d’Alacant, edición 0.9.2. Versión electrónica

<https://rua.ua.es/dspace/bitstream/10045/53085/6/librett-2016-v0.9.2-es.pdf>>

Hutchins, W. John; Harold L. Somers. (1992) *An Introduction to Machine Translation*. London: Academic Press.

Koehn, Phillip. (2009) "A process study of computer-aided translation." *Machine Translation* 23:4, pp. 241-263. Versión electrónica
<<https://link.springer.com/article/10.1007/s10590-010-9076-3>>

López Ciruelos, Andrés. (2003) "Una defensa crítica de las memorias de traducción" *Panace@* 4:12, pp. 180-182.

Oliver, Antoni; Joaquim Moré & Salvador Climent (coord.). (2007) *Traducción y tecnologías* Barcelona: UOC.

Ortega, John E.; Felipe Sánchez Martínez; Mikel L. Forcada. (2014) "Using any machine translation source for fuzzy-match repair in a computer-aided translation setting." *Proceedings of AMTA* -:1, pp.42-53. Versión electrónica
<https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_ResearchTrack_final.pdf>

Screen, Benjamin. (2016) "What does Translation Memory do to translation? The effect of Translation Memory output on specific aspects of the translation process." *Translation & Interpreting* 8:1, pp. 1-17. Versión electrónica <<http://www.trans-int.org/index.php/transint/article/view/506/242>>

Screen, Benjamin. (2017) "Productivity and quality when editing machine translation and translation memory outputs: An empirical analysis of English to welsh translation." *Studia Celtica Posnaniensia* 2:1, pp. 119-142. Versión electrónica
<https://www.researchgate.net/publication/319606846_Productivity_and_quality_when_editing_machine_translation_and_translation_memory_outputs_an_empirical_analysis_of_English_to_Welsh_translation>

Vázquez, Lucía Morado. (2013) “Comparing forum data post-editing performance using translation memory and Machine Translation output: a pilot study.” En: Sharon O’Brien; Michel Simard & Lucia Specia (eds.) 2013. Proceedings of the MT summit XIV workshop on post-editing technology and practice. Nice EAMT, 249-256. Versión electrónica
<https://www.researchgate.net/publication/272442095_Comparing_forum_data_postediting_performance_using_translation_memory_and_machine_translation_output_a_pilot_study>

10. Anexo

Apéndice de siglas	
FMS	<i>Fuzzy Match Score</i> (porcentaje de coincidencia)
LM	Lengua meta
LO	Lengua origen
MT	Memoria de traducción
TA	Traducción automática
TAO	Traducción asistida por ordenador
TMX	Translation Memory eXchange (Es un estándar de XML sirve para el intercambio de memorias de traducción)
UDMTA	Uso directo de la memoria de traducción automática
UDMTP	Uso directo de la memoria de traducción profesional
UIMTA	Uso inverso de la memoria de traducción automática
UIMTP	Uso inverso de la memoria de traducción profesional
UT	Unidades de traducción

Símbolos	
ED	Distancia de edición: número mínimo de inserciones, borrados o sustituciones de una palabra para obtener s a partir de s' .
L_1	Lengua que más domina un traductor profesional; normalmente hacia la que suele traducir.
L_2	Segunda lengua que más domina un traductor profesional; normalmente desde la que suele traducir
$ s $	longitud del segmento s
$ s' $	longitud del segmento s'
S	Memoria de traducción profesional directa
s	Segmentos en lengua LO
S'	Encargo de traducción (LO)
s'	segmentos de S' (LO)
T	Corpus en LM
t	Segmentos del corpus T
T'	Encargo de traducción obtenido (LM)
t'	Segmento del encargo de traducción (LM)
Z	Corpus en LO (<i>francés</i>) traducido con TA inversa de T
z	Segmentos en LO traducidos con TA inversa de t
V_2	Velocidad 2 (calculada en palabra/seg)
V_3	Velocidad 3 (calculada en palabra/seg) solo para los segmentos que sí ofrecían coincidencia en las MT
V_4	Velocidad 4 (calculada en palabras/seg) solo para los segmentos que no ofrecían coincidencia en las MT

Agradecimientos

Quiero agradecer a los participantes, mis amigos y compañeros de carrera, por haber dedicado parte de su tiempo a que se haya podido realizar este trabajo: Sofia Pina Pamies, Marina Soria Gómez, Nacho López Ucha y Ana Belén Sánchez Jiménez; también a Adrián por enseñarme sus conocimientos en estadística que me han ayudado a tratar de forma más eficiente los resultados obtenidos.